H-France Review Vol. 23 (February 2023), No. 22

Nicholas D. Paige, *Technologies of the Novel. Quantitative Data and the Evolution of Literary Systems*, Cambridge University Press, 2022. 287 pp. Index, figures, technologies section, and bibliography. $33.95 U.S. (pb.) ISBN 9781108890861.

Review by Christof Schöch, Universität Trier.

One of the arguments commonly used to motivate the use of computational methods in literary studies is that, to avoid telling literary history based only on a small selection of purportedly representative works, we would need to read more books than is humanly feasible. The solution proposed by the emerging field of Computational Literary Studies has been to design, digitize and analyse large collections of literary texts using quantitative methods such as stylometry, topic modeling, or network analysis.[1] The author of *Technologies of the Novel* proposes a different approach, much indebted to the early Franco Moretti of *Graphs, Maps and Trees*.[2] The author's aim is to provide "a largely descriptive account of a cultural system's behaviour" (p. 7), namely the French novel between 1600 and 1830. To achieve this, the author has collected, through shallow but systematic reading, a broad range of mostly formal features describing each item in a large collection of novels. In *Technologies of the Novel*, the author explores the evolution of and the interactions between these formal features over the course of time.

Before considering specific aspects of the book, a brief overview of the contents is in order. Chapter one tackles the issue of what motivates the recurring "truth posture" of eighteenth-century fiction, that is the phenomenon where author figures, fictional editors or narrators pretend to present found, authentic documents rather than advertising the fictional status of their discourse. The author investigates the interactions of this "truth posture" with other features like subject matter and narrative perspective, with the aim of understanding its "origins, diffusion, and eventual disappearance" (p. 19). Chapter two traces the rise and fall of what the author calls the "Aristotelian novel," a type of novel whose protagonist is modeled after a well-known historical or contemporary figure. Chapter three is concerned with various strategies of integrating insets, or embedded narratives, into the main body of the novel. Chapter four analyses the prevalence of explicit generic subtitles like *histoire* and *nouvelle* in comparison with unsubtitled novels. Relating subtitling to novel length, for example, the author shows that the *nouvelles* of the later seventeenth century are indeed shorter than the *histoires* of the earlier seventeenth century, but longer than the later *histoires*. Chapter five is similarly concerned with the evolution of *roman* and *nouvelle*; however, it investigates them independently of their generic subtitles and based, rather, on a number of formal characteristics. Chapter six is concerned with first-person narration, primarily the memoir novel, and their "documenticity" signals. The signaling of novels that are purportedly authentic documents comes and goes, but it appears that forms initially

associated with such signaling then continue their development while dropping the documenticity signals. Chapter seven moves on to trace the development of the French epistolary novel, distinguishing (based on both thematic and formal aspects) an unusually long period of thematic and formal experimentation up until around 1740 and a comparatively modest rise and peak of the recognizable polyphonic, sentimental epistolary novel between 1760 and 1800. Chapter eight, in turn, deals with the rise of the third-person novel in the late eighteenth century and specifically investigates whether it can be understood as the return to an earlier form of third-person narration or constitutes a new type, considering three key aspects: length, chapterization, and incipits. Chapter nine compares the development of the French novel, detailed in the previous eight chapters, to the development of the British novel. The comparison is illuminating, not only with respect to Ian Watt's infamous "rise of the novel," but also in terms of a number of similarities and differences in the development of the British and French epistolary novel.[3] Chapter ten concludes the book with a reflection on the underlying principle of the dynamics of the novel observed in the previous chapters and attempts to understand these dynamics primarily in terms of technological evolution.

All of the analyses presented in *Technologies of the Novel* are based on a dataset that is described in the annex to the book. The dataset mainly consists of several tables and is made freely available by the author.[4] Apart from a number of novels excluded for reasons detailed in the annex, the table on the French novel contains 1310 novels that have been retained for reading, annotation and analysis. A second list contains 460 British novels for the period from 1700 to 1830. The annex to the book explains the sampling procedure used to reduce the overall production of novels over 230 years to manageable proportions. The author first sampled one or several years from each decade, then collected all the available books from these years for inspection. This means that, for the decades starting with the 1770s, all data comes from just one or very few individual years, rather than from a larger number of years spread out across the decade. While the annex acknowledges that any sampling strategy implies a certain margin of error, such considerations play only a very limited role in the body of the book.

For each of the novels retained in the sample, the author then established a host of descriptive features, among them their year of first publication; their length (in words); their narrative perspective (e.g. first-person vs. third-person); their generic subtitle (e.g., *roman* or *nouvelle*); their mode of segmentation (e.g. into chapters), whether or not they use inset narratives (and which kind); their subject matter (contemporary or historical); their protagonists (known or unknown figures); their type of incipit (e.g. description of character vs. location); their formal type (e.g., epistolary vs. memoir); or their "truth posture" (e.g., whether they are openly advertised as fictional or, on the contrary, purportedly consist of found documents). To one degree or another, all of these categories are of course defined through simplification and abstraction. However, the author not only carefully discusses the design of each category. He also makes clear that a certain degree of abstraction is a necessary condition of comparability and quantification, a deliberate reduction of individuality driving the emergence of other kinds of knowledge related to prevalence and correlation.

The book's argumentation is then primarily based on the detailed analysis of the data available in spreadsheet form and presented to the readers in the form of more than 120 visualizations. In some cases, the author is interested in the development over time of certain isolated features. More often than not, however, several such features are combined to define a certain type of novel, whose development is then traced through the rising or falling numbers of relevant cases

in each decade. These types of novels are understood not in the sense of subgenres (thereby avoiding the pitfalls of genre theory), but as "formally similar narrative artifacts" (p. 138) that can be understood as "technologies" (hence the title of the book).

Given the centrality of the visualizations for the books argument, it appears essential to also examine the visualization strategies used. Beyond the ubiquitous simple lineplots and barcharts, there are several instances where they are used with particular sophistication. A case in point is the visualization showing the disappearance of the earlier *roman* and its replacement by the more recent *nouvelle* over the course of the seventeenth and eighteenth centuries (fig. 5.2 and 5.10 in the book; see also Paige's fig. 8.14). The key innovation here is to use multiple alternative definitions of both types of novels, based on more or less strictly defined formal characteristics, and to visualize the resulting trends together in a series of (filled or unfilled) lineplots that are shaded from dark to light corresponding to stricter or looser definitions. The result is a visualization that is both functional and aesthetic. Rather than reducing categorical complexity, it embraces it, thus maintaining a great deal of nuance.

In several instances, however, the simpler lineplots and barchars, while apparently easy to read, can also be somewhat misleading. The decision to summarize data in decades or periods of twenty years and visualize them as a continuous line suggests a continuity of development that is projected onto, rather than supported by the data. Also, both lineplots and barcharts obscure most of the uncertainty that is in the data due to the sampling procedure. Finally, no statistical tests are used to quantify the degree of correlation between two features or groups of novels, or to check the statistical significance of the difference between the means of certain distributions, relying entirely on a visual inspection of the visualizations. This can best be shown by looking more closely at one of the visualizations, figure 3.3, that shows the proportion of novels with titled insets among all Type 1 inset novels.

Commenting on figure 3.3. (replicated below as closely as possible based on the dataset provided), the author writes that the titled inset, "apparently introduced by d'Urfé, eventually becomes a nearly inevitable trait of the inset novel, especially the Type 1 inset novel" (p. 70). And he adds: "As Type 1 inset novels become more popular, titling the insets also becomes more popular" (p. 70).
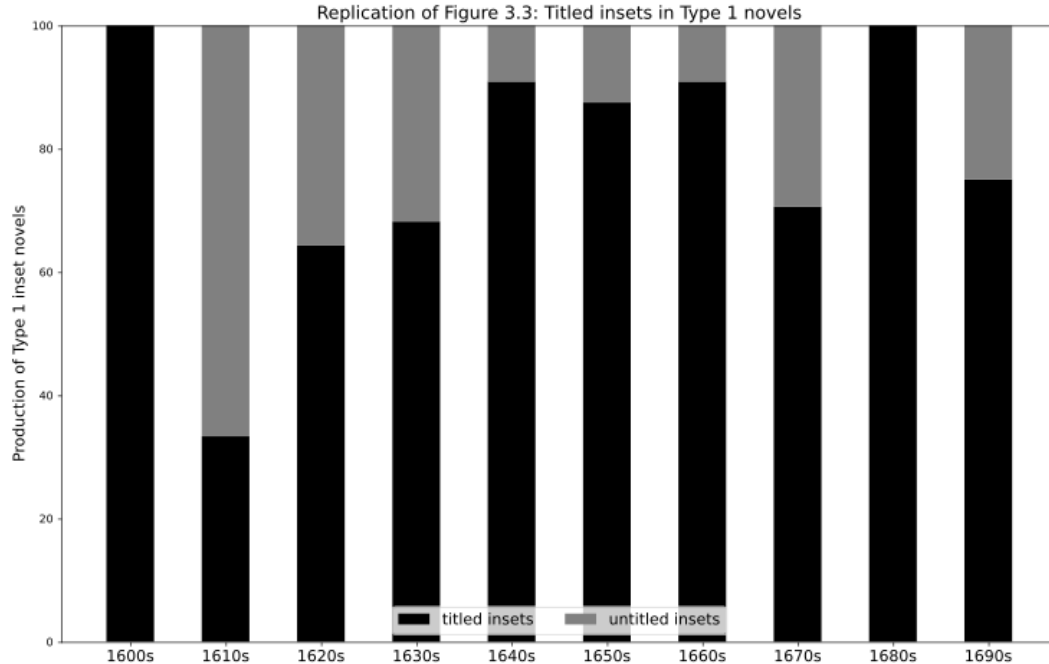
Replication of Figure 3.3 in Nicholas Paige, *Technologies of the Novel. Quantitative Data and the Evolution of Literary Systems* (Cambridge University Press, 2021), p. 71.
(reproduced with permission of Cambridge University Press through PLSclear)

Both of these observations are only weakly supported by the data, mostly because for such a rather specific question, only a small subset of the entire dataset will be relevant (in the case at hand, only 118 novels out of more than 1300). The author does note that for the 1600s, there is only one datapoint (d'Urfé's novel), so that the initial 100 percent proportion doesn't mean much. Four out of the ten decades (the first and last two), however, have less than five datapoints each. This means that contrary to the author's first observation suggesting a steady rise of the feature, any development or trend is in fact hard to support with the available data. They key observation we can make is that on average, over the entire century, about 75% of the Type 1 inset novels have titled insets. The second observation is, similarly, not borne out by the data: The correlation between the total number of Type 1 inset novels, and the proportion of titled insets, in each decade, is in fact very weakly negative, if present at all (at Pearson's R of -0.174).

The following alternative visualization, based on exactly the same dataset, attempts to make the above observations and uncertainties somewhat more transparent.

Re-analysis of Figure 3.3: Titled insets in Type 1 novels
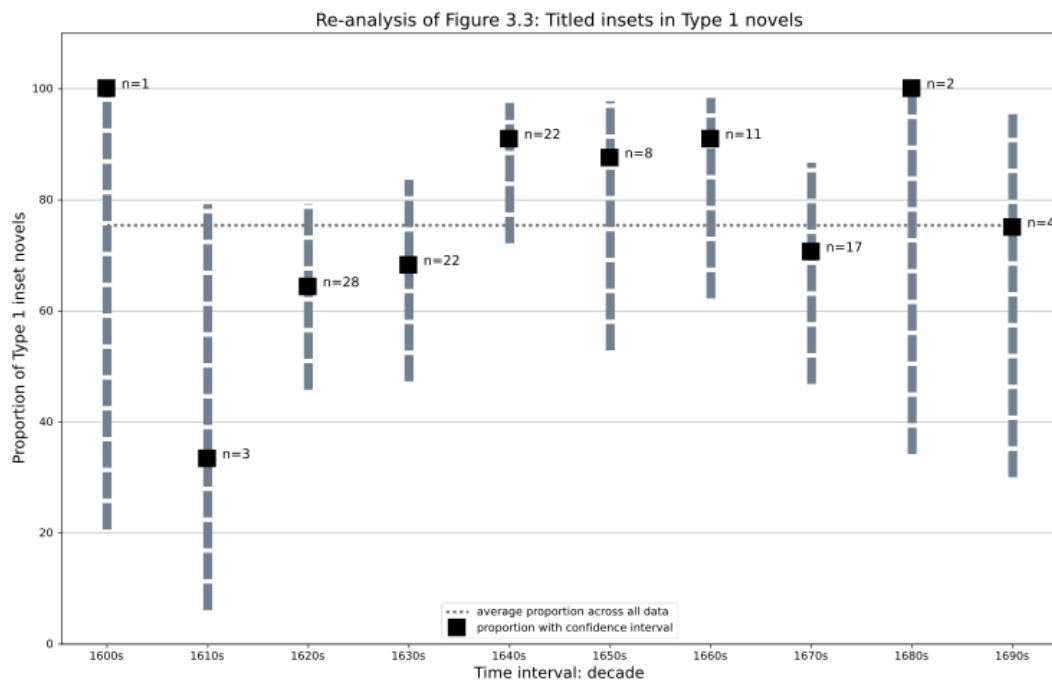
Instead of barcharts, the proportions of titled insets are shown as a single black square. To these observed proportions, two things are added: first, a number indicating the sample size for each decade; second, a confidence interval that shows, based on the sample size, the likely range of the true proportions, if all relevant novels had been taken into account. Simply put, the smaller the sample, the larger the uncertainty. If we discard the two earliest and the two latest decades that show a very high uncertainty, we are left with a slight up or maybe up-and-down movement. However, even then, the confidence intervals show substantial overlap, putting a doubt on the reality of this evolution.

The above analysis should not be misunderstood as a general indictment of the book's approach. In the case of figure 3.3, the small subset of the data is the main reason why the conclusions are built on uncertain ground. It is true that there are several other cases where trends and differences that appear to be self-evident in a lineplot become much more nuanced when the distribution of the data is taken into account (as in figure 4.4 on the median lengths of novels).[5] In most other analyses, however, the subset used is considerably larger. Therefore, the majority of the plots are solid bases for discussion. It is also important to note that only because the author has published his dataset, can this kind of checking and discussion even happen--a decision to be applauded.

The book's investigations of French literary history all have shown how formal features, content-based distinctions and contemporary labels appear to be connected in multiple ways. Not without the occasional polemic touch, the author contends that their interaction and joint development cannot be explained by factors external to the system of the novel itself (for instance, the absolutism of Louis XIV's reign causing the emergence of the French *nouvelle*, or the emerging bourgeois subjectivity explaining the rise of the epistolary novel). Similarly, he explicitly rejects

any attempts to see the history of the novel simply as one of progressive refinement, increasing complexity, and continuous innovation. Interestingly, he also and much more implicitly rejects attempts to use the domains of evolutionary biology or cultural evolution as a source of explanatory mechanisms, despite the fact that he readily uses the term "evolution" and defines it as "*constrained* change" (p. 6, emphasis in the original).

Rather, the author posits the novel and its various narrative forms as a set or system of "technological artifacts" and looks at the genre's development from the perspective of technological development, with W. Brian Arthur's *The Nature of Technology* as his guide. "The most basic position is that the literary forms I've isolated are evolving technological artifacts competing with others in the accomplishment of tasks that producers and consumers feel are important at any moment", he contends (p. 12). Discontinuous "redomaining" and incremental evolution, often with the latter following the former, are the two major explanatory mechanisms from W. Brian Arthur that the author mobilizes in this last chapter of the book that works somewhat as an afterthought to the book. To give just one example, the author explains the appearance of the memoir not so much as a gradual and progressive evolution from the earlier *nouvelle,* but as the relatively sudden integration of a form developed outside the novel system (the authentic historical memoirs) into the realm of fictional narrative that effectuates a radical change in the technology, before the new technology is then incrementally refined.

Before concluding, it may be appropriate to return to a question frequently levied against computational approaches to literary history: What can we learn, in this manner, that we did not know already or that we could not also learn through close reading? Simply put, Paige's book makes it very clear that the system-level view of 230 years of literary history, grounded in solid empirical observations, is not something that we have been able to obtain through close reading. More particularly, any observation of features correlating in their development over time, or of explicitly defined forms replacing each other in consecutive waves, is something we just cannot see quite as clearly using other methods. So what new knowledge has Paige found? Plenty of details, but also some interesting more general points. One among several such findings concerns the life cycle of a new type of novel. Such a form typically appears to start out in a prolonged phase of slow gestation that includes quite a bit of experimentation and heterogeneous production. The form then becomes more homogeneous at the same time as becoming more and more prevalent, until it reaches its peak while pushing one or several older forms out. The fortunes of the form then fade, while the next new form begins to rise, creating sometimes overlapping, sometimes distinct waves: from the (old) *roman* to the *nouvelle,* then to the *memoires,* then the *roman épistolaire,* and onwards to the (new) *roman.*

Ultimately, and despite an often rather simplifying visualization strategy, one can only be impressed with this highly original, thorough, and thought-provoking book. Anyone interested either in the history of the novel (whether French or not), or in current methods of Computational Literary Studies, should be reading *Technologies of the Novel.*

NOTES

[1] See, for instance, Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History* (Champaign: University of Illinois Press, 2013), Katherine Bode, *A World of Fiction: Digital Collections and the Future of Literary History* (Ann Arbor: University of Michigan Press, 2018), Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago, Ill: University of

Chicago Press, 2019) and Frank Fischer and Daniil Skorinkin, "Social Network Analysis in Russian Literary Studies," in D. Gritsenko, M. Wijermars, and M Kopotev, eds., *The Palgrave Handbook of Digital Russia Studies* (Palgrave Macmillan, 2021), pp. 517-536.

[2] Franco Moretti, *Graphs, Maps and Trees: Abstract Models for Literary History* (London: Verso Books, 2005).

[3] Ian Watt, *The Rise of the Novel: Studies in Defoe, Richardson and Fielding* (1957).

[4] See: Nicholas D. Paige, "Datasets and electronic graphs for Technologies of the Novel". Zenodo, July 2020. DOI: 10.5281/zenodo.3939065. The availability of this dataset follows an emerging best practice in the field and enables fellow researchers to repeat and build on the author's analyses, in the spirit of transparency, replicability and sustainability of research now summarized under the heading of Open Science.

[5] For more detailed analyses of these two cases and some more attempts at replication and re-analysis of Paige's dataset, see the following page: https://github.com/christofs/paige.

Christof Schöch
Universität Trier
schoech@uni-trier.de